Information Intelligence Available in **ONLINE** and **CD-ROM** editions. ISSN 0737-7770

ONLINE LIBRARIES AND MICROCOMPUTERS

George S. Machovec, Managing Editor
Head, Library Technology & Systems
Arizona State University, Tempe, AZ

NEWS & TRENDS

MARCH 1992
VOL.10 NO.3

WAIS: WIDE AREA INFORMATION SERVERS

"It's a jungle out there!" The Internet or "the Net", as it is commonly called, provides access to terabytes of information for thousands of nodes. This includes not only online library systems but massive amounts of information in data files which may be downloaded from computer centers around the world, ListServ conferences, commercial databases, governmental sources and the list goes on. In the last year, the definition of who are acceptable participants on the Internet has been broadened so that virtually any organization or person can qualify in one way or another. The entrance of many of the commercial information utilities (e.g. DIALOG, STN, OCLC, RLIN) on the Net is indicative of this change. The problem is no longer access, but how to know what is available and how to navigate in this environment.

Over the last few years many different directories have been compiled to assist in this navigation process. Numerous directories have been compiled for library systems alone -- the most famous probably being those by Billy Barron and another by Art St. George. These types of tools are very helpful and can be downloaded (using the FTP command) for no charge. However, this type of action already assumes a certain level of literacy in networking and even tools such as these really do not tell a user what kind of data may be found in each system. One project to help to solve this dilemma has been under development for the last couple of years is the **Wide Area Information Server** (WAIS, pronounced "ways").

History

The WAIS project began as an experimental venture between four companies: Thinking Machines Corporation (producer of massively parallel computers and information retrieval engines), Apple Computer, Dow Jones & Company, and KPMG Peat Marwick. The purpose of the project was to create an easy-to-use interface which could access many information servers regardless of location. In addition, the interface would not require that the user become familiar with all of the different systems and data could be delivered to the requester without regard to their origins.

From the user's perspective, there are a number of problems to solve. First one must identify and select information from databases from a very large pool of choices. Second, these databases may be on a variety of different systems and the user should not be required to know how to use each system. Third, there needs to be some way

more -

EDITORIAL BOARD

Joseph Becker, President, Becker and Hayes, Inc., Pacific Palisades, CA
James B. Dodd, Reference Department, Georgia Tech Libraries, Atlanta, GA
Miriam A. Drake, Dean/Director of Libraries, Georgia Inst. of Tech., Atlanta, GA
Allan D. Pratt, American Graduate School of Int'l Management, Glendale, AZ

T.M.

© 1992 - Information Intelligence Inc., P.O. Box 31098, Phoenix, AZ 85046 - (602) 996-2283 - Published monthly except July and August and mailed first class or overseas airmail. No part of this publication may be reproduced by any means without prior written permission of the publisher. Photocopying for internal use permitted only if appropriate fees paid to Copyright Clearance Center, Inc., or publisher. $6.25 fee applies to any copying of an issue whether complete or incomplete. Subscription rates: See last page of each issue for current rates. Abstracted and indexed in Library and Information Science Abstracts (LISA) and Referativnyi Zhurnal.

(WAIS - cont'd)
to download and organize the retrieved data so that one is not overwhelmed.

WAIS

To solve this set of problems, the wide area information server (WAIS) concept was developed. It is based on the client/server architecture and WAIS was designed using an extension of the Z39.50 NISO protocol. The basic standard was used for bibliographic retrieval but was extended to handle the needs of full-text retrieval, imaging and audio. This allows access to a diversity of servers and so that clients may access a wide variety of information, not just bibliographic.

To make the project successful beyond the experimental stage, there must be a critical mass of Z39.50 compliant servers which use the extended WAIS protocol. Thinking Machines has published the specifications for the protocol, which is available at no charge, although it comes with little support. It will be up to the information providers to develop the WAIS compatible server software to become a player in this market.

One of the most significant aspects of the WAIS project is the development of this extended Z39.50 open protocol, thus creating an ad hoc market where system developers want to become participants. Although this sounds like "pie in the sky", the reality is that there is a growing momentum of information providers on the Internet who are enthusiastically supporting the WAIS project and have or are developing WAIS compliant servers.

This protocol is hardware independent, thus defusing the complaint that applications must be developed under one brand of hardware or operating system. The key is interoperability through the WAIS protocol and not forcing all information providers to use one brand of hardware, software, or even search interface. The WAIS software handles the negotiations on each system.

Initially, the client computer software (for the end-user) was developed on an Apple Macintosh platform, but subsequent work has been done to port the WAIS client software to MS-DOS and UNIX machines. This helps ensure the long-term success of the project so that the client machines are not limited to one brand of equipment.

How Does it Work for the End User?

Interaction with the WAIS system occurs through the Question interface. This is a graphical user interface (GUI) which employs pull-down menus (Mac-like) and although result sets may look different, for example ASCII text will have a different looking display then some downloaded bit-mapped graphics, the user only needs to become familiar with the one interface to gather information from many different information servers.

To begin a session the user pulls down a query window and asks a question in a natural English-language style, thus not requiring the need to know special query languages. In the next step the user pulls down a menu identifying servers which will be queried for the appropriate information (the WAIS interface can also identify source systems if the user does not know what to select). After the appropriate information is retrieved from the remote servers headlines of materials are displayed in a window and the user may "point and click" on any relevant result to retrieve the information.

more -

©1992 Information Intelligence Inc

(WAIS - cont'd)

Behind the user interface a series of activities are in progress. The natural language English query is translated into the WAIS protocol. This is then transmitted over the network to each server, which then takes the natural language request and puts it into a structured search which operates on its own local information retrieval engine. Once a list of relevant answers has been retrieved they are then encoded into the WAIS protocol and sent back over the network to the client where the results appear. It can be seen from this technique that hardware and software independence is achieved since both the client and the server side of the system translate queries and results into the intermediate WAIS protocol.

Difficulties

Although the WAIS project solves many problems for the "digital researcher", a number of difficulties immediately become apparent:

- Because natural language is used by the user and then translated by a remote server into its own query language a number of difficulties could arise: the natural language query may be ambiguous, the translation into a local system query language will work better on some systems than others, a local system may or may not be able to handle some types of searches, and so on.

- Since the original search query was developed in a free natural language environment, it may have to be interpreted by the server into a fairly "low level" keyword search. This is because many of the advanced search features available on local online systems may not be known by the end-user or even be able to be entered into a WAIS natural language query.

- Although many information nodes on the Internet are "free" there are a growing number of commercial services that have connect-hour, hit-charges and full-text delivery fees. Methods and techniques will need to be developed to handle these costs in the WAIS model and users may need to identify up-front "free" systems. If systems are selected which have charges methods (such as 900 numbers, passwords or credit-card solutions), will need to be implemented. Obviously, information is not free and to have a critical mass of useful information, the commercial publishers, vendors and database producers need to be involved. Flexibility in both the WAIS design and forward looking thinking for database producers are needed.

- As there will be a growing number of servers and databases available through the WAIS system, the issue of relevance feedback becomes more important. Queries which are too broad (or narrow), improperly constructed or not appropriate, may generate huge retrieval sets (or null sets at the other extreme). This is already a problem when searching large stand-alone databases where a person is interacting in real-time. The problems may explode many times over in the WAIS system and automatic techniques for limiting (or broadening) results will need to be developed. The current WAIS model extracts keywords from natural language queries and does not extract semantic information, this will need to be refined to provide more useful retrieval when searching large databases from multiple sources.

more -

©1992 Information Intelligence Inc

(WAIS - cont'd)

- If a large number of sources are defined by the user either local or remote, results may take some time before being returned and sorted. To avoid communication delays in the Internet, especially when downloading large search sets, queries are executed and results returned at night. This means that users will have a delay in viewing results which is fine in some situations but not acceptable in other cases. If one needs immediate results it may still mean a direct log-on to an information source to get real-time results.

- One of the features being developed in the WAIS environment is to develop tools to support current awareness services (similar to SDIs which have traditionally been employed on many online hosts). In this concept, a user's profile can be periodically run against the identified WAIS servers so that new relevant documents can be downloaded into the user's result windows. To do this the WAIS client software must be developed to clearly identify what are new sources, techniques must be improved to quickly view stored documents, and there is the need for larger computer screens to easily view large amounts of text.

How to Find the WAIS Servers

As a growing body of information nodes become compatible with the WAIS system, it is clearly not practical for any one user to keep track of what is available. To solve this problem, Thinking Machines is maintaining a Directory of Servers which contains indexed text-based descriptions of all known servers.

In this model, if the user does not know where to go for information, the query would first be presented to the Directory of Servers which would reply with a list of possible databases and servers on which it was available. The user would then formulate the query and identify the possible sources for the actual execution of the request.

The Directory of Servers is also intended as a central source for database producers or system servers to advertise new databases. The producer can provide the textual descriptions for the product as well as information on how to contact the server(s) on which the product is mounted.

Conclusion

The WAIS system is one of the most exciting developments on the Net today. It is becoming more than just a "good idea" since life on the Net is not only exciting but brutal. Today, information literacy must involve an in-depth knowledge of not only what electronic information is available but also how to get there. Libraries must have on their list of priorities the need to educate their users on network use and also should be involved with putting information resources on the national networks. Anything less than this may eventually result in patrons meeting their information needs without us. [GSM]

Selected Bibliography

Stein, Richard Marlon. "Browsing Through Terabytes", Byte 16(5):157+ (May 1991).

Kahle, Brewster and Art Medlar. "An Information System for Corporate Users: Wide Area Information Servers", Online 15(5):56-60 (September 1991).

more -

©1992 Information Intelligence Inc

(WAIS - cont'd)

Davis, Franklin et. al. "WAIS Interface Protocol Prototype Functional Specification", Thinking Machines. Available from Franklin Davis on the Internet by contacting fad@think.com or by contacting Brewster Kahle at brewster@think.com.

FCLA SIGNS AGREEMENT TO LOAD IAC DATABASES

The Florida Center for Library Automation (FCLA) has signed an agreement with Information Access Company (IAC) to load two major databases, the Expanded Academic Index and Business Index.

FCLA runs the NOTIS system for the state universities which include Florida A&M University, Florida State University, Florida Atlantic University, University of Central Florida, University of Florida, University of North Florida, University of South Florida and University of West Florida. The two IAC databases will be available at all of these sites and will represent more than 3.5 million IAC records. The FCLA system currently has more than 7 million MARC records for the consortium.

For more information contact: Information Access Company, 362 Lakeside Drive, Foster City, CA 94404. Telephone: (800) 227-8431.

CARL SYSTEMS & BNA ANNOUNCES INTERFACES

CARL Systems, Inc. and Blackwell North America (BNA) have announced an agreement in January 1992 to develop a series of interfaces using the Internet communications network. The BNA services that will be available to CARL users will include:

- access to BNA's NTO (new titles online) database enriched with abstracts and tables of contents

- electronic ordering of materials from both BNA and B.H. Blackwell Ltd (BHB)

- electronic distribution of LC MARC cataloging enhanced with tables of contents and processed through Blackwell authority control

- a unique interface with the upcoming CARL System authority control module providing Blackwell's retrospective processing, a current and ongoing service providing immediate updating of new records, and a notification service that will maintain CARL System files in constant accordance with the most current practice of the Library of Congress.

The link will also allow Blackwell's approval customers to receive electronic transmission of weekly packing list/invoice data through CARL System's acquisition system.

For more information contact: CARL Systems Inc., 3801 E. Florida, Ave., Bldg. D - Suite 300, Denver, CO 80210. Telephone: (303) 758-3030.

©1992 Information Intelligence Inc

ELSEVIER SCIENCE PUBLISHERS TO DISTRIBUTE SOME JOURNALS ELECTRONICALLY

Elsevier Science Publishers is initiating a program called **TULIP** (The University Licensing Program) in which it will distributed 35 of its materials science and engineering journals on magnetic tape. TULIP is intended to begin as a 3-year (1992-1994) project in 12-16 academic institutions who would like to acquire the 35 titles on magnetic tape for loading into a local information system.

The academic institutions would allow the libraries to subscribe to the tapes on a flat-fee basis for institutional use and would also establish a rate structure if the library would like to provide electronic access to articles to corporate users. To participate in the project, Elsevier is requiring that organizations must subscribe to at least half of the titles in paper copy and are members of the Coalition for Networked Information.

Not every participating library will locally load the articles. Several key nodes will be selected and they will provide access to the other participants over the Internet. Elsevier will be gathering use data in this experimental project although they have indicated that they will not publish detailed use data about specific articles or titles, although more general information will be disclosed. For more information contact: North-Holland, Elsevier Science Publishers BV, P.O. Box 1991 - 1000 BZ Amsterdam, The Netherlands. In the U.S. and Canada contact: Elsevier Science Publishing, P.O. Box 1663, Grand Central Station, New York, NY 10163.

VTLS TO DEVELOP ITS SYSTEM TO RUN UNDER UNIX

VTLS, Inc. has announced that it has begun development of a version of its integrated online library system to run under the Unix operating system, which is becoming a popular operating system for many computer manufacturers. VTLS has a design team in place for the project and its initial work will be to port VTLS to the Hewlett-Packard HP9000 Unix machine. A beta test version is scheduled for testing during the third quarter of 1992.

The VTLS software currently runs on HP and IBM computers, but the popularity of the Unix operating system has prompted the company to offer this popular option. The Unix version will use a relational database management system and no functionality of the application will be lost. VTLS will also offer a migration path for its existing customers who would like to change from the HP (MPE) or IBM (VM) operating system to Unix.

More information is available from: Vinod Chachra, President, VTLS Inc., 1800 Kraft Drive, Blacksburg, VA 24060. Telephone: (703) 231-3605. Fax: (703) 231-3648.

PBS AND DATA TREK SIGN AGREEMENT

Personal Bibliographic Software and Data Trek, Inc. have signed an agreement in which Data Trek will become a distributor of PBS's database management program, Pro-Cite, and Biblio-Link USMARC, which supports the transfer of MARC records from Data Trek into Pro-Cite. In this arrangement, users of Data Trek's Professional and Manager Series Cataloging modules can put Data Trek MARC records automatically into a Pro-Cite database.

For more information contact: Data Trek Inc., 5838 Edison Place, Carlsbad, CA 92008. Telephone: (800) 876-5484, (619) 431-8400. or PBS Inc., P.O. Box 4250, Ann Arbor, MI 48106. Telephone: (313) 996-1580. Fax: (313) 996-4672.

©1992 Information Intelligence Inc